Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

The structure and domain organization of *Escherichia coli* isocitrate lyase

K. L. Britton,[a] I. S. B. Abeysinghe,[b] P. J. Baker,[a] V. Barynin,[a] P. Diehl,[c] S. J. Langridge,[a] B. A. McFadden,[c] S. E. Sedelnikova,[a] T. J. Stillman,[a] K. Weeradechapon[a] and D. W. Rice[a]*

[a]Krebs Institute for Biomolecular Research, Department of Molecular Biology and Biotechnology, The University of Sheffield, Sheffield S10 2TN, England, [b]Biochemistry Division, Tea Research Institute, Talawakelle, Sri Lanka, and [c]Department of Biochemistry and Biophysics, Washington State University, Pullman, WA 99164-4660, USA

Correspondence e-mail: d.rice@sheffield.ac.uk

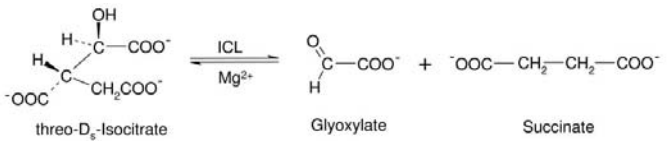
Received 8 December 2000
Accepted 25 May 2001

PDB Reference: isocitrate lyase, 1igw.

Enzymes of the glyoxylate-bypass pathway are potential targets for the control of many human diseases caused by such pathogens as *Mycobacteria* and *Leishmania*. Isocitrate lyase catalyses the first committed step in this pathway and the structure of this tetrameric enzyme from *Escherichia coli* has been determined at 2.1 Å resolution. *E. coli* isocitrate lyase, like the enzyme from other prokaryotes, is located in the cytoplasm, whereas in plants, protozoa, algae and fungi this enzyme is found localized in glyoxysomes. Comparison of the structure of the prokaryotic isocitrate lyase with that from the eukaryote *Aspergillus nidulans* reveals a different domain structure following the deletion of approximately 100 residues from the larger eukaryotic enzyme. Despite this, the active sites of the prokaryotic and eukaryotic enzymes are very closely related, including the apparent disorder of two equivalent segments of the protein that are known to be involved in a conformational change as part of the enzyme's catalytic cycle.

1. Introduction

Isocitrate lyase (ICL) catalyses the first unique step in the glyoxylate cycle, a metabolic bypass which was first proposed by Kornberg & Krebs (1957) to account for microbial growth on two-carbon compounds and the conversion of lipids to carbohydrates in fatty seedlings. In this cycle, ICL and malate synthase (MS) function sequentially to convert isocitrate to succinate plus glyoxylate and glyoxylate plus acetyl-CoA to malate, respectively. Together, catalysis by these enzymes ensures the bypass of two oxidative steps of the tricarboxylate acid cycle (TCA) in the synthesis of succinate. Thus, the glyoxylate cycle is carbon-conserving and ensures an adequate supply of TCA-cycle intermediates for biosynthetic purposes when cells convert lipids to carbohydrates.



In bacteria, enzymes for the metabolism of two-carbon compounds including ICL and MS are found predominantly in the cytoplasm. In that compartment isocitrate is partitioned between the energy-yielding TCA cycle and the carbon-conserving glyoxylate cycle, with each branch initiated through catalysis by isocitrate dehydrogenase or ICL, respectively. In contrast, in eukaryotes ICL is found in peroxisomes often termed glyoxysomes. Biochemical studies have established that the most common quaternary structure

for ICL is a tetramer (Gainey *et al.*, 1992). However, analysis of the available amino-acid sequences reveals that the subunit molecular weight varies between 48 kDa for the prokaryotic enzymes and 67 kDa for the enzyme from eukaryotes. Comparison of the sequences revealed that this difference is largely associated with an insertion of approximately 100 residues towards the C-terminus of the enzyme in the eukaryotic ICLs; it has been proposed that these residues may play a role in the differential targeting of the enzymes to their different cellular compartments. The recent structure determination of the ICL from *A. nidulans* (Britton *et al.*, 2000) has revealed that the subunit structure of this enzyme is based on two domains, one of which resembles a TIM barrel. The second domain is seen to be composed mainly of α-helices formed from a block of approximately 100 additional residues which are missing from the prokaryotic enzymes. Analysis of the *A. nidulans* ICL structure revealed two segments of disordered residues which lie close to the active site and which contain a number of conserved residues. The recent structure determination of *Mycobacterium tuberculosis* ICL has confirmed that these regions are important in catalysis.

The two key enzymes of the glyoxylate cycle, ICL and MS, are widely distributed among microorganisms including *Mycobacteria*, *Pseudomonas aeruginosa*, a major contributor to pathogenesis in cystic fibrosis (Shimamoto & Berk, 1980) and five species of *Leishmania*, some of which are insidious human parasites. In this latter protozoan, a functional glyoxylate cycle (Simon *et al.*, 1978) offers an inviting target for control of human infections as this pathway is not thought to operate in man. A number of isocitrate lyase-directed inhibitors are known including nitropropionate (Schloss & Cleland, 1982) and itaconate (McFadden & Purohit, 1977), which function to control growth of many organisms utilizing the glyoxylate cycle (Vanni *et al.*, 1990). In order to contribute to a fuller understanding of the structure–activity relationships in this enzyme family, we have carried out the structure determination of the ICL from *E. coli*. In this paper, we report the structure of a Ala219Cys mutant of this enzyme which retains full biological activity and whose analysis allows a detailed comparison of the prokaryotic and eukaryotic enzymes.

2. Materials and methods

2.1. Data collection and processing

Purified protein was prepared from an overexpressing *E. coli* strain and crystals of Ala219Cys *E. coli* ICL were grown in 0.1 *M* HEPES, 5 m*M* magnesium chloride pH 7.0 using the hanging-drop method of vapour diffusion with 20–22%(*w*/*v*) methyl-ethylpolyethylene glycol (MEPEG) 2000 as a precipitant in the presence of 5 m*M* EMTS. These crystals belong to space group $P3_2$ and have unit-cell parameters $a = b = 88.7$, $c = 199.4$ Å with a tetramer in the asymmetric unit. X-ray diffraction data were collected from a single crystal at room temperature on a MAR image-plate detector on station 9.5 at the CCLRC Daresbury synchrotron to a resolution of 2.1 Å.

Table 1
X-ray data collection and refinement statistics.

	All data	Outer shell
Resolution (Å)	33.3–2.1	2.15–2.1
No. of observations	188246	13435
No. of unique reflections	96329	6893
Completeness (%)	94.1	91.3
Multiplicity	2.0	1.9
$R_{\rm merge}$† (%)	5.1	21.6
Data with $I/\sigma(I) > 3$ (%)	86.9	67.3
Refinement resolution limits (Å)	15.0–2.1	
No. of reflections used in refinement	88762	
No. of reflections for $R_{\rm free}$	4666	
R value ($R_{\rm free}$)	0.184 (0.235)	0.234 (0.297)
R.m.s. deviation bond distance (Å)	0.007	
R.m.s. deviation bond angle (°)	1.6	
R.m.s. deviation peptide planarity (Å)	0.010	
R.m.s. deviation aromatic planarity (Å)	0.002	
Total non-H atoms	13432	
Water sites	767	
Mercury sites	21	
Magnesium ion sites	4	
Pyruvate atoms	24	
Mean r.m.s. B factors for main chain (Å)	1.3	
Mean r.m.s. B factors for side chain (Å)	2.8	
Mean B factors (Å^2)		
All atoms	31	
Main chain	29	
Side chain	33	
Water molecules	34	
Mercury	37	
Magnesium ions	19	
Pyruvate atoms	21	
Ramachandran statistics for each monomer‡ (%)		
Most favoured regions	94.2, 92.2, 92.9, 93.5	
Additional allowed regions	5.5, 7.0, 6.5, 5.6	
Generously allowed regions	0.3, 0.8, 0.3, 0.3	
Disallowed regions§	0.0, 0.0, 0.3 (H197), 0.6 (E159, F322)	

† $R_{\rm merge} = \sum_{hkl} |I_i - I_m| / \sum_{hkl} I_m$, where I_i and I_m are the observed intensity and the mean intensity of related reflections, respectively. ‡ The quality of the structures of each of the four subunits (*A*, *B*, *C* and *D*) in the asymmetric unit were assessed using *PROCHECK* (Laskowski *et al.*, 1993) and statistics are given for each monomer independently. § Residues occurring in the disallowed region have been identified.

These data were processed using the *MOSFLM* suite of programs (Leslie, 1992) and the resulting data were merged and scaled using the programs *ROTAVATA* and *AGROVATA* (Collaborative Computational Project, Number 4, 1994). Data-collection statistics are supplied in Table 1.

2.2. Structure determination

The structure was solved by a combination of molecular replacement and averaging/phase extension based upon a model of the *A. nidulans* ICL structure (Britton *et al.*, 2000) modified after close inspection of the aligned sequences of the two enzymes. This resulted in the deletion of residues 268–362 which form an additional domain expected to be absent in the prokaryotic enzyme and also the deletion of regions from the model of the *A. nidulans* ICL structure which correspond to the extended N-terminus and a loop region (1–8 and 154–158, respectively). A cross-rotation function was calculated using the program *POLARRFN* (Collaborative Computational Project, Number 4, 1994) using data in the resolution range 10–5 Å with a radius of integration of 30 Å and gave a clear

solution of peak height 6σ. The program *TFFC* (Collaborative Computational Project, Number 4, 1994) was then used to calculate the translation function, which produced a solution with a peak height of 19σ. An initial electron-density map based on the molecular-replacement solution was calculated to 2.1 Å resolution and procedures of averaging over the four subunits in the asymmetric unit and solvent flattening were applied, gradually increasing the resolution from 8.0 to 2.1 Å in three increments, using the program *DM* (Cowtan, 1994). Molecular masks for the molecule were generated using the program *NCSMASK* (Collaborative Computational Project, Number 4, 1994). A total of 720 cycles of solvent flattening and molecular averaging reduced the free *R* factor from 0.580 to 0.309.

2.3. Construction of the model and refinement

In the resultant electron-density map clear density could be found for much of the polypeptide chain for each of the subunits in the asymmetric unit. An initial model comprising 380 of the expected 434 amino-acid residues of the *E. coli* ICL could be built with confidence using the graphics program *FRODO* (Jones, 1985) for one of the subunits, with the other subunits of the tetramer being generated by non-crystallographic symmetry. The model was refined using the program *REFMAC* (Murshudov *et al.*, 1997). 21 Hg atoms arising from the modification of the enzyme by EMTS during the crystallization were added to the model at an early stage of refinement. Solvent molecules were also added using the program *ARPP* (Lamzin & Wilson, 1993) and individual *B* factors refined for all atoms. In the latter stages of refinement all four monomers were treated independently. In the final model solvent molecules were rejected if their *B* factors were above 60 Å^2 after refinement and the occupancies of the Hg atoms were adjusted to maintain an average *B* factor close to that of the protein. Iterative cycles of rebuilding and refinement gave a final crystallographic *R* factor for the model of 0.184 (free *R* factor 0.235) for all data (88 762 reflections) in the 15–2.1 Å resolution range. The overall average *B* value for the tetramer is 31 Å^2 (29 Å^2 for main-chain atoms). Table 1 gives details of the refinement statistics and model quality.

3. Results and discussion

3.1. Crystallization of Ala219Cys *E. coli* ICL

The Ala219Cys mutant of *E. coli* ICL was one of a series of cysteine mutants initially constructed to aid determination of the structure of wild-type *E. coli* ICL by isomorphous replacement. This mutant enzyme shows activity of 15.3 U mg^{-1} (where U is the amount of enzyme which will catalyze the transformation of 1 µmol of substrate per minute), compared with the wild-type ICL of 22.0 U mg^{-1} (Conder *et al.*, 1988), and crystallizes in the presence of ethyl mercury thiosalicylate (EMTS) under conditions that are closely related to those previously described (Abeysinghe *et al.*, 1991). These crystals belong to the trigonal space group $P3_2$, with a tetramer in the asymmetric unit. The recent solution of the structure of *A. nidulans* ICL (Britton *et al.*, 2000) made an isomorphous replacement study unnecessary as molecular replacement could be used to solve the structure. The crystals of the Ala219Cys mutant diffract to beyond 2.1 Å resolution compared with diffraction to only 2.5 Å observed with crystals of the wild-type enzyme. Given the similar catalytic properties of these two enzymes, we decided to determine the Ala219Cys mutant ICL structure in preference.

3.2. Structure solution

In the final map of the A219C mutant of *E. coli* ICL the electron density is of high quality for most of the protein atoms. Of the four subunits (labelled *A*, *B*, *C* and *D*), breaks in the electron density occur only in a few locations which are in general at equivalent regions of the structure in each subunit. Thus, in each subunit there is no density corresponding to the N-terminal methionine or residues 418 to the C-terminus (434). In subunit *C* the remainder of the polypeptide chain can be seen, albeit with weaker electron density for residues Glu141–Arg146 and His197–Met198. In the remaining three subunits the density in two segments of the chain, Lys193–Gly199 and Ser319–Ala334 and a number of flanking residues, is either disordered or somewhat weaker than that observed in subunit *C*, whilst for subunits *A* and *D* the density is also weak between Glu141 and Arg146. In subunit *C*, six Hg atoms can be located, whereas for subunits *A*, *B* and *D* only five can be seen. This is consistent with the additional disordering in the latter subunits of Cys195. Thus, the final model for the tetramer comprises the coordinates for 1628 of the expected 1736 residues (*A* subunit, 396 residues; *B* subunit, 411 residues; *C* subunit, 416 residues and *D* subunit, 405 residues) together with 21 Hg atoms, an overall total of 767 solvent molecules and four active-site magnesium ions. For subunits *B*, *C* and *D* the density for Pro320 is clear and consistent with a *cis*-proline residue (Fig. 1). In addition, an unexplained

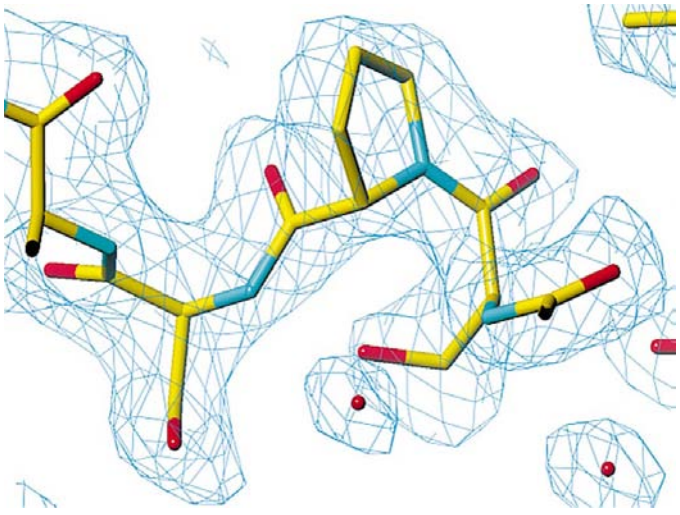


Figure 1
The electron density for subunit *C* in *E. coli* ICL around the loop containing the *cis* proline (Pro320) in the final $(2|F_o| - |F_c|)\ \alpha_{calc}$ map calculated using the refined model at 2.1 Å.

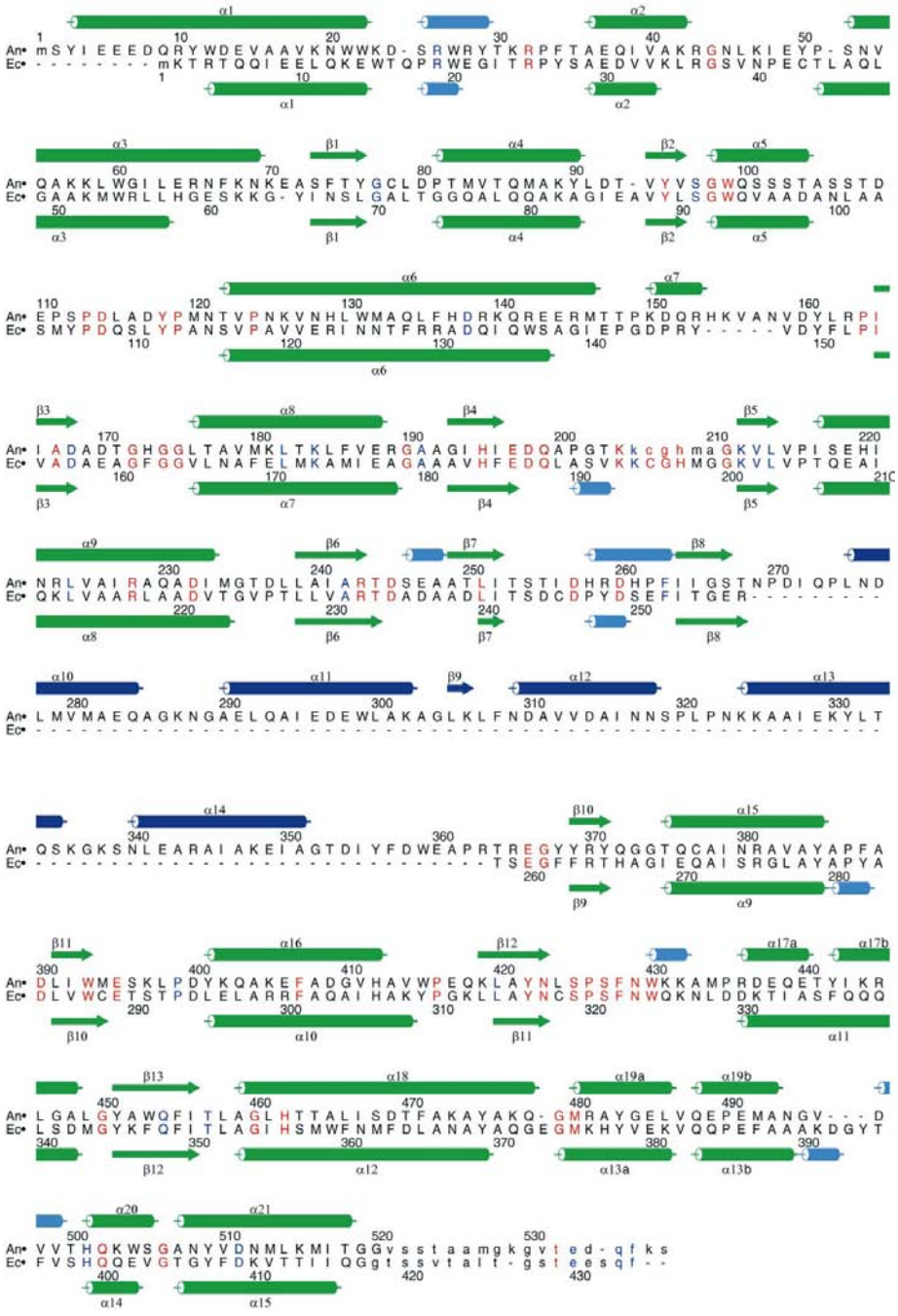


Figure 2
Structure-based sequence alignment of ICL from *E. coli* (ACEA_*E. coli*) and from *A. nidulans* (ACEA_EMENI). The alignment was derived by consideration of the sequences from 23 ICL sequences (detailed below); residues conserved over all these ICLs are shown in red, with those conserved across 22 of the 23 highlighted in blue. Lower case has been used to denote the residues which are disordered in the most complete subunits of the two enzyme structures and are therefore absent from the model. The sequence numbering of every tenth residue in the two sequences is also shown. Positions of deletions are indicated by -. The secondary-structural elements of the three-dimensional structures are shown above and below the sequence alignment with helices shown as cylinders and strands represented by arrows. 3_{10}-helices are coloured pale blue and are not labelled. The secondary-structural elements corresponding to domain II of the *A. nidulans* ICL are coloured dark blue. The other ICL database sequences used to produce this alignment and not shown here are: ACEA_BRANA, ACEA_CORGL, Q10663, AF013216, ACEA_RHOFA, Q39577, ACEA_MYCLE, ACEA_NEUCR, O13439, ACEA_CANTR, ACEA_YARLI, ACEA_YEAST, ACEB_YEAST, ACEA_PINTA, ACEA_RICCO, ACEA_GOSHI, ACEA_CUCSA, ACEA_LYCES, P93110, ACE1_SOYBN, ACE2_SOYBN. The figure was drawn using the program *ALSCRIPT* (Barton, 1993).

feature can be seen in the map in the region of all active sites whose shape and size can be fitted with molecules such as pyruvate/glyoxylate/oxalate. An unknown molecule, built as pyruvate, has been included in the model to account for this density. All side chains were fitted, but the average temperature factors for 51 of the total of 1628 residues in the asymmetric unit were at least 70 Å^2. Ramachandran plots showed each subunit having at least 92% of the residues in the most favoured regions (Table 1).

3.3. Overall fold and tertiary structure of *E. coli* ICL

The ICL subunit comprises a single domain of approximate dimensions 40 × 50 × 60 Å, which is formed from a total of 15 α-helices and 12 β-strands together with a small number of 3_{10}-helices and helical turns. These elements of regular secondary structure together comprise 58% of the polypeptide chain. The secondary-structure assignments, determined using the criteria of Kabsch & Sander (1983), are given in Fig. 2 and a schematic representation of the structure of a single subunit is given in Fig. 3.

The N-terminal region of the polypeptide chain starts with the residue Lys2, the first observable residue in the electron-density map. This leads immediately into three consecutive helices (α1, α2 and α3) that lie on the periphery of the subunit. The structure then starts to fold into a central parallel β-sheet of eight strands (β1, β2, β3, β4, β6, β10, β11 and β12) alternating generally with one or more α-helices to give an overall fold of a $(\beta/\alpha)_7\beta$ barrel whose fold is therefore closely related to the TIM barrel seen in many other proteins (Chothia, 1988). Helix α3 lies at the N-terminal end of the TIM barrel and partially shields this portion of the structure from the solvent. Beyond the last strand in the TIM barrel (β12) the chain folds into a series of α-helices (α12, α13, α14 and α15) which form an extension leading towards the C-terminus. However, the polypeptide chain ends at residue 417 out of the

expected total of 434 residues. As in *A. nidulans* ICL, the helices α12 and α13 from a twofold-related subunit take the place of the eighth helix usually found flanking the barrel.

3.3.1. Comparison of the four subunits of the tetramer. Though the four subunits are complete to varying degrees, the tertiary structures of the monomers are essentially identical,

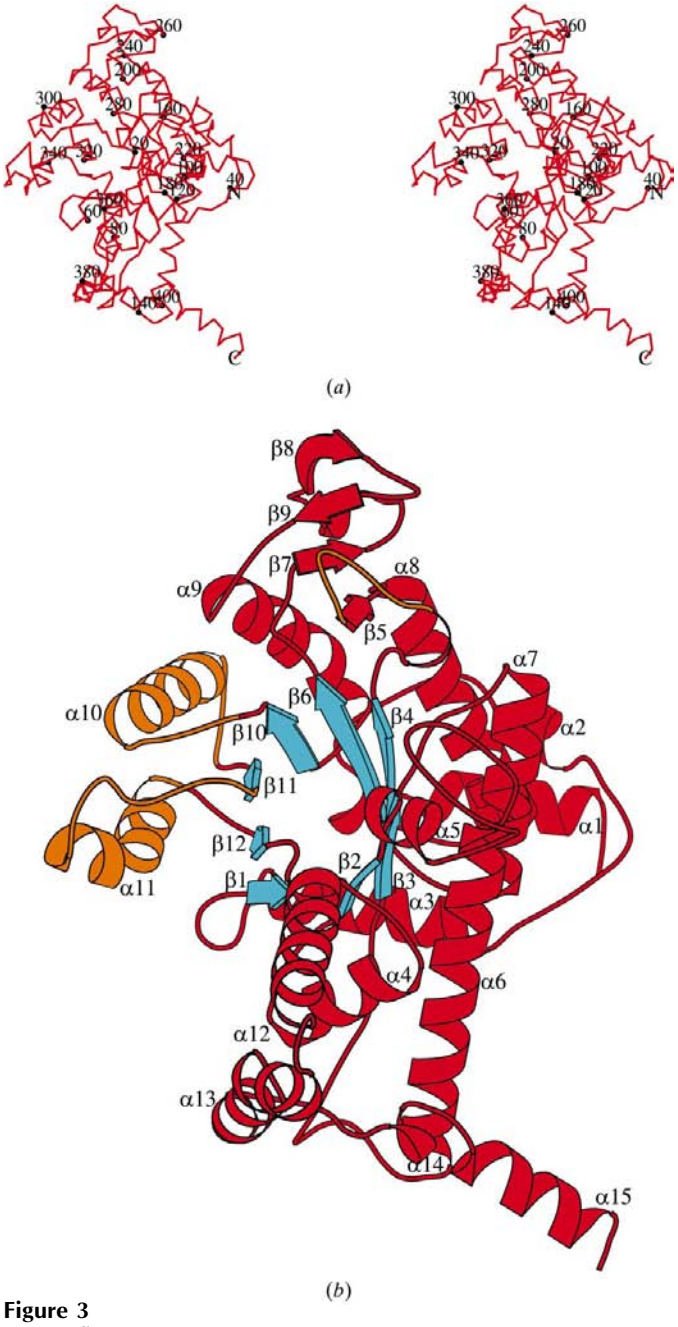


Figure 3
(*a*) A C^{α} chain trace of the *C* subunit of *E. coli* ICL with every 20th residue indicated by a black dot. This figure was produced using the program *MIDASPLUS* (Ferrin *et al.*, 1988). (*b*) A single subunit of *E. coli* ICL with the strands and helices labelled. The strands of the TIM barrel are shown in cyan and the helices in red, except for three regions where small deviations occur between the different subunits (orange). Viewed from the top in an anticlockwise rotation, these regions involve residues 192–200, 290–311 and 318–344. The C-terminal residues which are not observed in any subunit are not shown. Figure prepared using *MOLSCRIPT* (Kraulis, 1991)

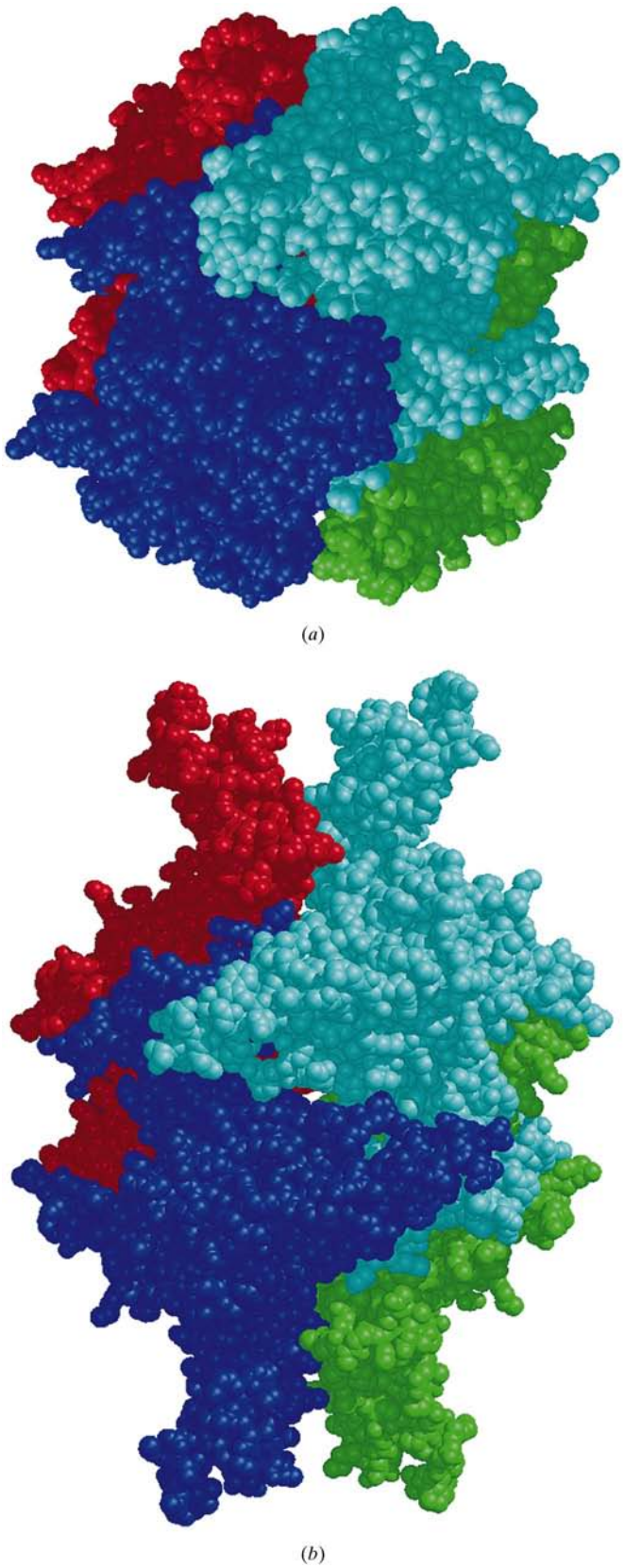


Figure 4
A space-filling representation for the ICL tetramers of (*a*) *E. coli* and (*b*) *A. nidulans*, with each subunit individually shaded highlighting the loss of the head domain in the simpler prokaryotic enzyme. This diagram was produced using the program *MIDASPLUS* (Ferrin *et al.*, 1988; Huang *et al.*, 1991).

with superpositions of the *A*, *B* and *D* subunits onto that of the most complete *C* subunit revealing root-mean-square deviations of 0.25–0.46 Å. Small deviations between the four subunits are found consistently in three places, all of which lie at the C-terminal end of the barrel. Two of these include the segments of the chain that are the sites of variable disorder in the subunits (Fig. 3*b*).

The presence of EMTS in the crystallization buffer resulted in the binding of several Hg atoms which have been included in the enzyme model. Of the six cysteine residues per subunit in the mutant enzyme, all are exposed to the solvent and all appear to be accompanied by an Hg atom except for the site adjacent to Cys195 of subunits *A*, *B* and *D*. This residue is only visible in the electron-density map for subunit *C* and consequently the location of any bound mercury at the disordered cysteine positions of the other three subunits is also unclear. The Hg atoms are bound to each of the remaining cysteine side chains at equivalent positions.

3.3.2. Subunit interfaces and quaternary structure. In the biological tetramer the subunits are related by 222 symmetry to give an almost spherically shaped molecule with an approximate radius of 40 Å. Accessible surface-area calculations (Lee & Richards, 1971) show that a single subunit of this ICL has an accessible surface area of 20 550 Å^2 and on forming a tetramer 36% (7350 Å^2) of this surface is buried (all values taken to the nearest 50 Å^2). The contact surface is composed of three different dimer interfaces about the three orthogonal molecular twofold axes which involve burying 800, 2250 and 4300 Å^2 of the subunit solvent-accessible surface area, respectively.

3.4. Structural similarities of *E. coli* ICL to other proteins and implications for function

3.4.1. Structural similarity and domain organization of prokaryotic and eukaryotic ICLs. *E. coli* ICL shares a 40% sequence identity with its eukaryotic counterpart from the fungus *A. nidulans*. The sequence of *E. coli* ICL is significantly shorter than that of the *A. nidulans* enzyme since the 95 residues corresponding to 269–363 in *A. nidulans* ICL are not present. These residues interrupt the fold of the β-barrel motif immediately after β8 to form a second peripheral head domain in the *A. nidulans* ICL structure. In the *E. coli* enzyme, therefore, this second domain is absent (Fig. 4), although only a minor rearrangement of the local structure appears to be necessary to accommodate this structurally. The function of the second domain in *A. nidulans* ICL is uncertain, but previous work has suggested a possible role in targeting this enzyme to glyoxysomes (Matsuoka & McFadden, 1988). Comparison of the TIM-barrel domains of these two structures (using subunit *C* of the *E. coli* enzyme) shows that they are closely related except where insertions and deletions give rise to local changes, most notably at the N-terminus. Small differences appear to affect the relative lengths of the secondary-structure elements and may arise owing to the fact that the *A. nidulans* model was determined at lower resolution. All the above differences lead to relatively small changes which do not affect the central TIM-barrel motif where the active site of the enzyme has been shown to be located (Britton *et al.*, 2000). Overall, these structural differences between the *E. coli* and *A. nidulans* ICLs do not affect assembly of the tetramer. The subunit interfaces in the *E. coli* enzyme are entirely analogous to those found in the *A. nidulans* ICL (Britton *et al.*, 2000) both in terms of the buried surface areas and chemical character of the residues present at each interface.

3.4.2. Analysis of the active site in *E. coli* and *A. nidulans* ICLs. Early work to identify the glyoxylate-binding site in the structure of *A. nidulans* ICL was hampered by the apparent binding to the enzyme of an as yet unidentified ligand for which electron density was seen in the original MIR map. This was thought to be a buffer component, inhibitor or substrate possibly associated with the enzyme-bound metal ion and carried through during purification. Subsequently, soaking experiments on crystals grown from protein which had been exhaustively dialysed before crystallization allowed the glyoxylate-binding site to be located unambiguously by difference Fourier methods. Glyoxylate binds at the C-terminal end of the TIM-barrel motif, consistent with the general location of the active site in other enzymes with this fold (Britton *et al.*, 2000). Examination of the electron-density map for the *E. coli* enzyme revealed a similar unexplained discrete electron-density feature associated with the same piece of protein structure and occupying the same position as the glyoxylate in the *A. nidulans* ICL.

In the *A. nidulans* enzyme, the glyoxylate moiety can be seen to interact with the conserved residues at positions Ser97 (side-chain hydroxyl group), Arg243 (side-chain

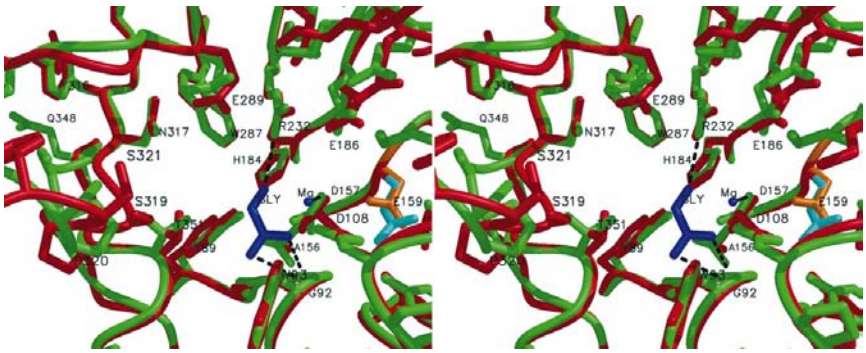


Figure 5
Stereoview of the top of the TIM-barrel region of *E. coli* (red) and *A. nidulans* (green) ICL subunits, highlighting the side-chain positions of the conserved residues. In both structures those residues which are conserved in at least 22 of the 23 ICL sequences are shown. In addition, the two equivalent residues Glu159 in *E. coli* (cyan) and Asp170 in *A. nidulans* (orange) are highlighted. The location of the magnesium ion and the glyoxylate in *A. nidulans* ICL are shown in blue together with the interactions they make to the protein. This figure was drawn using the program *MIDASPLUS* (Ferrin *et al.*, 1988).

guanidinyl group) and Trp99 (amino N atom) as well as the magnesium ion. In *E. coli* ICL, the relative positions of the equivalent residues and the metal ion are very similar, suggesting that this part of the active site is strongly conserved. In the structure reported here the unidentified electron-density feature has been modelled as a pyruvate ion.

A wider comparison of the two ICL structures shows that most of the conserved residues in the putative active-site pocket have identical or nearly identical conformations, including three of the four carboxyl groups found close to the divalent metal ion (Asp108, Asp157 and Glu186) and nearby conserved serine (Ser91), tyrosine (Tyr89), glycine (Gly92), arginine (Arg232) and histidine (His184) residues. One difference appears to be the conservative substitution of Asp170 in the *A. nidulans* enzyme by Glu159 in the *E. coli* ICL. In both enzymes, these equivalent residues lie near the divalent metal ion and the sequence change leads to a somewhat different position for the carboxyl group but with subtle changes in the main chain which result in Glu159 lying in a generously allowed region of the Ramachandran plot (Fig. 5). Another difference between the two enzymes close to the active site can be seen in the vicinity of residues 319–321. The sequence over this region is strongly conserved in ICL and includes two serines whose side-chain hydroxyl groups have been implicated as being important in stabilizing the dioxyanion of the aci-carboxylate of succinate (Ko *et al.*, 1992; Rehman & McFadden, 1997). Whilst originally built as a *trans* proline, as observed in the structure of *A. nidulans* ICL, the electron density for Pro320 found in subunits *B*, *C* and *D* of the *E. coli* structure suggests it is in the *cis* conformation (Fig. 1). Re-examination of the *A. nidulans* model confirmed the *trans* assignment in this structure. The reason for this difference is currently unclear. However, we note that the *E. coli* enzyme has been modified by mercury at the adjacent cysteine (Cys318), whereas no such modification has been carried out in the *A. nidulans* enzyme.

3.4.3. Conformational changes in catalysis by ICL. In the structures of both *E. coli* and *A. nidulans* ICL the location of an equivalent segment of the polypeptide chain following the final helix to the C-terminus is disordered (residues 418–434 and 521–538 of the two enzymes, respectively). In *A. nidulans* ICL the presence of a second disordered region corresponding to a loop containing a number of totally conserved residues across the entire ICL family (residues 205–210 including Lys205, Cys206, Gly207, His208) is mirrored in three of the subunits of the *E. coli* enzyme (residues 194–199). However, in the *C* subunit of *E. coli* ICL the path of the polypeptide chain can be discerned, although the density does weaken around residue His197. This loop can be seen to extend above the C-terminal end of the barrel close to the active site of the enzyme. There are no apparent crystal contacts which might account for the ordering of the region of the structure in subunit *C*. In addition, for the remaining three subunits disorder is also seen in an adjacent region which includes a conserved SPS motif (319–321) which is thought to be important in the recognition of succinate. Disorder of this region has not been noted previously.

Very recently, the structure determination of the ICL from *M. tuberculosis* has been reported in the absence of ligands and in complex with two inhibitors (Sharma *et al.*, 2000). Comparison of the structure of the *M. tuberculosis* free enzyme with that of *E. coli* ICL reveals a root-mean-square deviation of 0.75 Å over 379 equivalent C^{α} atoms and shows that this enzyme also lacks the second domain found in the *A. nidulans* enzyme. The structure of the free *M. tuberculosis*

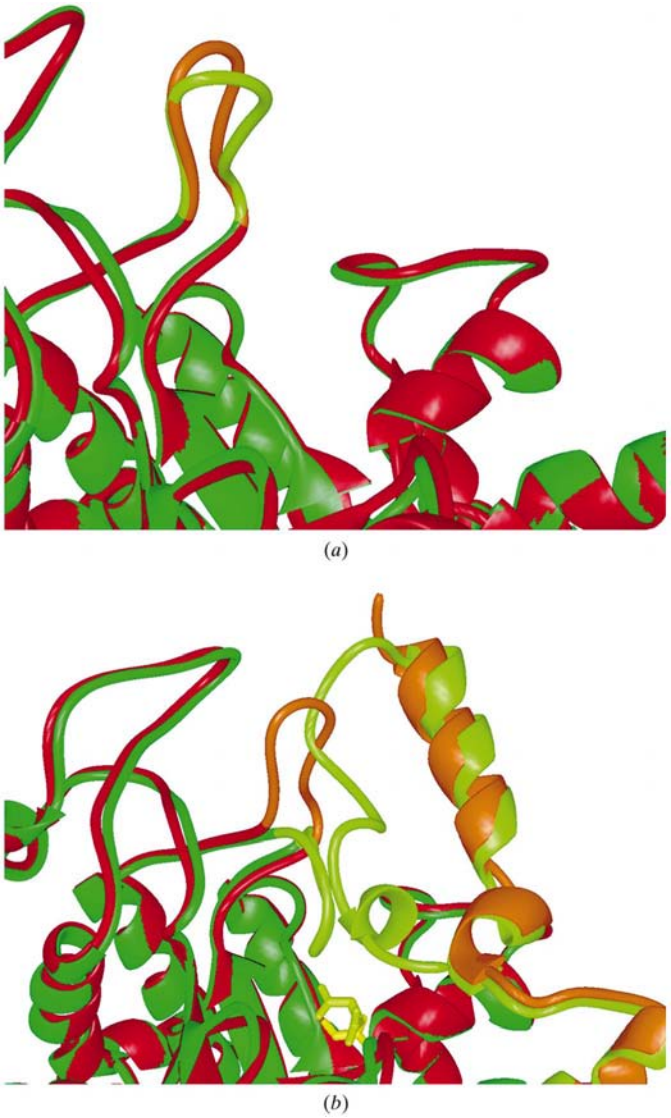


Figure 6
(*a*) A close-up of the structural overlap between *E. coli* mutant ICL *C* subunit (red) and the *M. tuberculosis* 'free' enzyme (green). Drawn using the program *MIDASPLUS* (Ferrin *et al.*, 1988). The position of the loop carrying the active-site cysteine (Cys195 in *E. coli* ICL) is shown in orange, whilst the equivalent loop region in the *M. tuberculosis* enzyme is highlighted in lime. (*b*) A close-up of the structural overlap between *E. coli* mutant ICL *C* subunit (red) and the *M. tuberculosis* enzyme (green) complexed with glyoxylate (yellow) and nitropropionate (lime). Drawn using the program *MIDASPLUS* (Ferrin *et al.*, 1988). The position of the loop carrying the active-site cysteine in *E. coli* ICL (Cys195) together with the C-terminal residues of the neighbouring subunit are shown in orange. The equivalent loop region and the location of the remaining C-terminal residues in the *M. tuberculosis* enzyme are highlighted in lime green.

ICL reveals disorder at the C-terminus. In contrast to the disorder of the highly conserved loop, residues 205–210 in *A. nidulans* ICL, the equivalent loop in *M. tuberculosis* ICL is ordered and lies in a similar position to that in the *C* subunit of the *E. coli* ICL structure (Fig. 6*a*). In the two structures of the *M. tuberculosis* ICL complexed with the inhibitors 3-bromopyruvate and 3-nitropropionate, the enzyme appears to undergo a conformational change which results in the ordering of the C-terminal residues and a dramatic movement of the cysteine-containing loop to lie on top of the active site. In the inhibitor complexes, the loop appears to act as a 'lid', opening to allow substrates and products to exchange with the solvent and closing in a manner which excludes bulk solvent from the active site during catalysis. The residues at the C-terminus fold up to form a short helix which packs against the cysteine-containing loop of a symmetry-related subunit (Fig. 6*b*). In *E. coli* ICL it has been suggested that Cys195 is the base for the enzyme reaction (Britton *et al.*, 2000; Sharma *et al.*, 2000). This residue is located on the loop region which is subject to conformational changes and in its open conformation would be somewhat remote from the active-site region. Loop closure, once the substrates have been bound, would result in this cysteine being more suitably placed to abstract the proton.

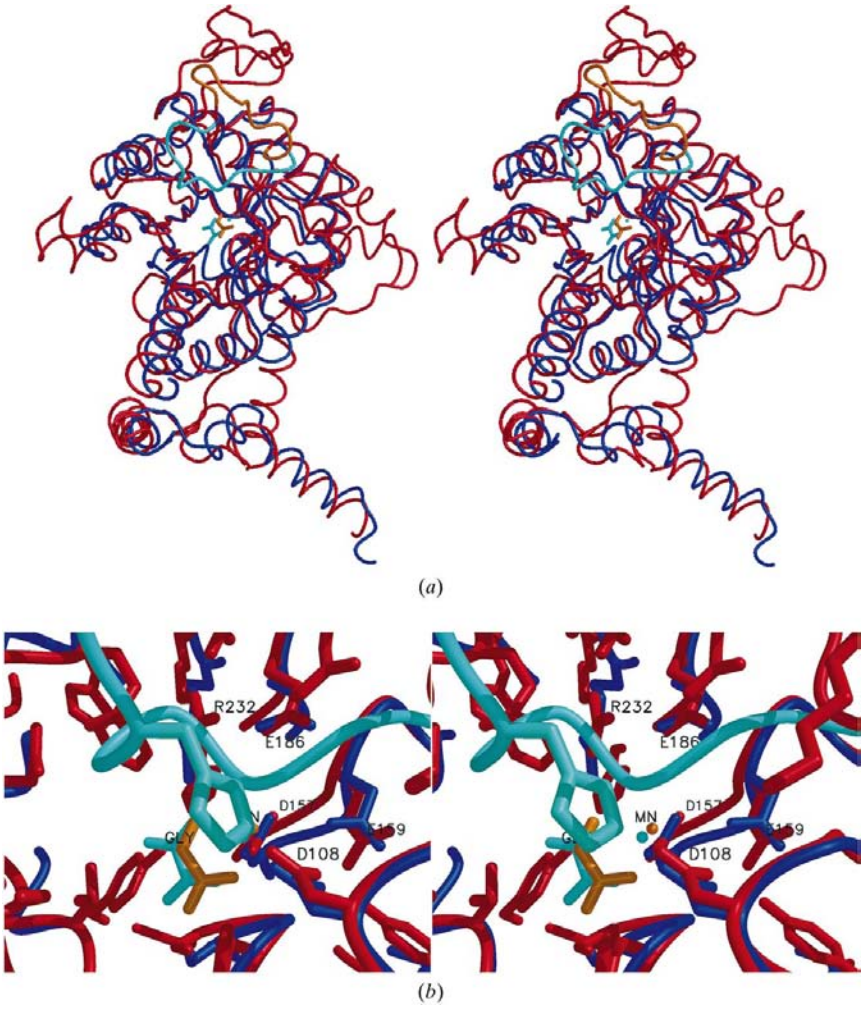


Figure 7
Stereoviews of the structural overlap between *E. coli* ICL *C* subunit (red) and the PEP mutase from *Mytilus edulis* (blue). Drawn using the program *MIDASPLUS* (Ferrin *et al.*, 1988). (*a*) The position of the loop carrying the active-site cysteine (C195) in *E. coli* ICL is shown in orange, which contrasts with the equivalent loop region in PEP mutase highlighted in cyan. The location of the bound magnesium ion and the inhibitor oxalate (cyan) in the PEP mutase are also shown which closely match the positions of the manganese ion and glyoxylate substrate found in *A. nidulans* ICL (orange). (*b*) Stereoview close-up of the overlap between *E. coli* ICL and PEP mutase in the region of the active site calculated following superposition of residues conserved across all sequences of both enzymes. The residues conserved in at least 22 of the 23 ICL sequences known are highlighted in red. The side chains highlighted in blue are those of PEP mutase residues which are identical to those found in ICL. The location of the closed active-site loop in the PEP mutase (cyan) can be seen, lying over the binding sites for the metal ion and inhibitor, and includes a conserved histidine also found in ICL.

3.4.4. Catalysis by ICL, a member of the enolase superfamily. ICL has previously been reported as a member of the enolase superfamily (Britton *et al.*, 2000; Babbitt *et al.*, 1996). Members of this family share the mechanistic similarity of involving an inital catalytic step of the removal of a proton α to a carboxyl group by a base on the enzyme.

A number of enzymes with the TIM-barrel fold and belonging to the enolase superfamily have also been shown to possess loops which move during catalysis. These include mandelate racemase (Neidhart *et al.*, 1991) and muconate lactonizing enzyme (Helin *et al.*, 1995). In both of the latter enzymes the loops are provided by small and equivalent N-terminal subdomains which sit adjacent to the TIM-barrel domain. Yet despite this similarity, these loops appear to perform different roles during the catalytic cycles of the two enzymes. In mandelate racemase the probable role of this loop is to supply the remaining hydrophobic residues which complete the pocket around which the phenyl substituent of the substrate binds. In contrast, the suggested role of the loop region in muconate lactonizing enzyme is that of excluding water from the active site during catalysis. Comparison of the locations of these mobile loops reveals that they do not interrupt the barrel at equivalent positions to that of the mobile loop in ICL.

Whilst ICL is clearly related in structure and mechanism to other members of the enolase superfamily, the closest resemblance revealed following the determination of the structure of the *A. nidulans* ICL is to the *Mytilus edulis* phosphoenolpyruvate (PEP) mutase (Huang *et al.*, 1999; PDB code 1pym). A detailed comparison between the structures of *E. coli* ICL and *M. edulis* PEP mutase shows that they are

more closely related than are *A. nidulans* ICL and PEP mutase, with both enzymes lacking the second domain seen in the *A. nidulans* enzyme. Thus, overall 221 C^{α} atoms can be equivalenced between *E. coli* ICL and PEP mutase, with an r.m.s. of 1.51 Å (Fig. 7*a*). This closer structural similarity is reflected in a marginally higher overall sequence identity between these two enzymes (18.4 *versus* 16.6% for *E. coli* and *A. nidulans* ICL *versus* PEP mutase, respectively).

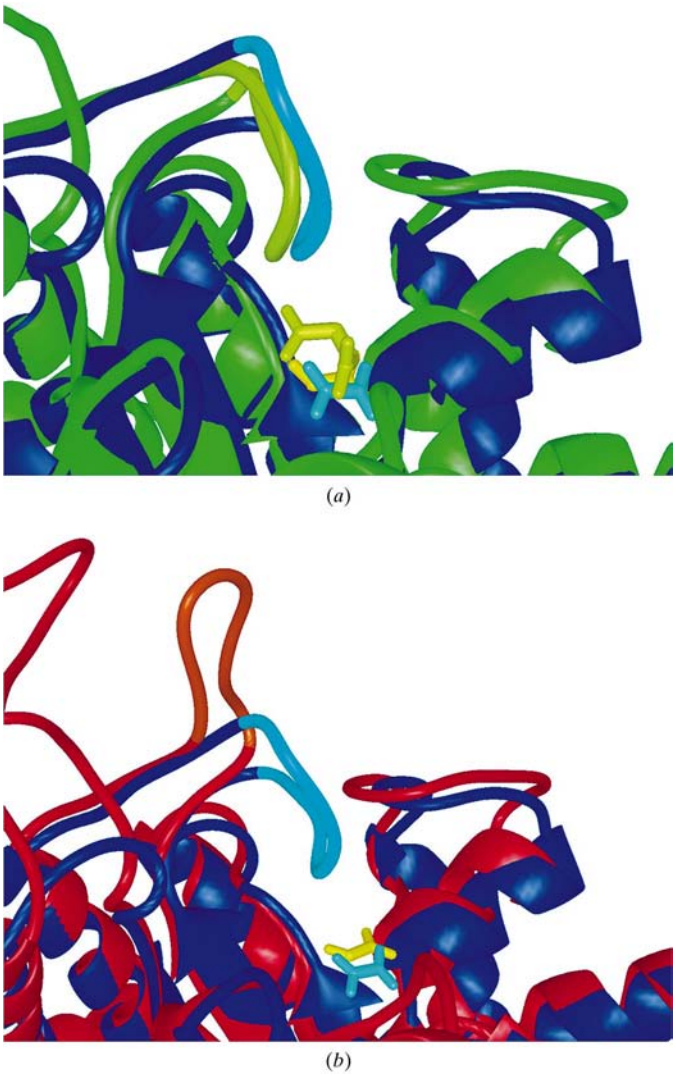


Figure 8
The same view and style as shown in Fig. 6. (*a*) Structural overlap, based on the superposition of 221 C^{α} atoms, between *Mycobacterium tuberculosis* ICL (green) and the *Mytilus edulis* PEP mutase (blue). The position of the loop carrying the active-site cysteine (Cys195 in *E. coli* ICL) is shown in lime green, whilst the equivalent loop region in the *M. edulis* PEP mutase is highlighted in cyan. The location of the inhibitor oxalate (cyan) in the PEP mutase is also shown, which closely matches the position of the bound glyoxylate substrate in the *M. tuberculosis* ICL (yellow). In addition, the location of an ICL inhibitor, nitropropionate, is seen (lime green) as determined from the *M. tuberculosis* ICL complex. (*b*) Structural overlap, based on the superposition of 221 C^{α} atoms, between *E. coli* mutant ICL *C* subunit (red) and the *M. edulis* PEP mutase (blue). The position of loop carrying the active-site cysteine (Cys195) in *E. coli* ICL is shown in orange, whilst the equivalent loop region in the *M. edulis* PEP mutase is highlighted in cyan. The location of the inhibitor oxalate (cyan) in the PEP mutase is also shown, which closely matches the position of the glyoxylate substrate found in *A. nidulans* ICL (yellow).

PEP mutase also possesses a loop which closes off the active site from the solvent in the structure of the enzyme complexed with the inhibitor oxalate (Fig. 7*b*). This has led Huang *et al.* (1999) to speculate that a conformational change is necessary to allow access to the active site for the enzyme's substrate. Comparison of ICL and PEP mutase reveal that the conformationally mobile regions are spatially equivalent, lying between strands β4 and β5 in the barrel. Analysis of the structures of the *M. tuberculosis* ICL–inhibitor complex with the PEP mutase–inhibitor complex indicates the positions adopted by the loops to be very similar (Fig. 8*a*). Comparison of the ICL and PEP mutase sequences for this loop reveals that this stretch of the polypeptide chain is highly conserved in each enzyme family. However, there appears to be very little sequence identity over this region between the two enzymes. Comparison of the relative positions of the two loops in the *E. coli* ICL and the *M. edulis* PEP mutase shows that they occupy different locations (Fig. 8*b*). The overlap between these structures reveals that the loops first deviate at residue Gln188 in *E. coli* ICL (199 in *A. nidulans* ICL and 116 in PEP mutase) and reunite again at residue Gly200 (211 in *A. nidulans* ICL and 129 in PEP mutase), suggesting that these residues may provide the flex points for the conformational rearrangement. Sharma *et al.* (2000) suggest that in *M. tuberculosis* ICL the conformational change is triggered by a 2.5 Å movement of the enzyme-bound Mg^{2+} ion and that this is in turn induced by the binding of succinate. This is consistent with the observation that the binding of glyoxylate to the crystals of *A. nidulans* ICL does not appear to lead to closure (Britton *et al.*, 2000). In contrast, in PEP mutase the binding of oxalate which occupies the equivalent region of the active site to glyoxylate in ICL (Fig. 8*b*) promotes the loop to close, suggesting that the trigger for the conformational change in the two enzymes is different.

We thank the support staff of the Synchrotron Radiation Source at the Daresbury Laboratory for assistance with station alignment. This work was supported by grants from The Wellcome Trust, BBSRC and New Energy and Industrial Development Organization. SIBA was supported by a grant from the Tea Research Institute of Sri Lanka. KW is supported by a scholarship from the Royal Thai Government. The Krebs Institute is a designated BBSRC Biomolecular Sciences Centre and the Krebs Institute structural studies group is a member of the North of England Structural Biology Centre (NESBIC).

References

Abeysinghe, S. I. B., Baker, P. J., Rice, D. W., Rodgers, H. F., Stillman, T. J., Ko, Y. H., McFadden, B. A. & Nimmo, H. G. (1991). *J. Mol. Biol.* **220**, 13–16.
Babbitt, P. C., Hasson, M. S., Wedekind, J. E., Palmer, R. J., Barrett, W. C., Reed, G. H., Rayment, I., Ringe, D., Kenyon, G. L. & Gerlt, J. A. (1996). *Biochemistry*, **35**, 16489–16501.
Barton, G. J. (1993). *Protein Eng.* **6**, 37–40.

Britton, K. L., Langridge, S. J., Baker, P. J., Weeradechapon, K., Sedelnikova, S. E., DeLucas, J. R., Rice, D. W. & Turner, G. (2000). *Structure*, **8**, 349–362.
Chothia, C. (1988). *Nature (London)*, **333**, 598–599.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Conder, M. J., Ko, Y. H. & McFadden, B. A. (1988). *Prep. Biochem.* **18**, 431–442.
Cowtan, K. (1994). *Jnt CCP4/ESF–EACBM Newsl. Protein Crystallogr.* **31**, 34–38.
Ferrin, T. E., Huang, C. C., Jarvis, L. E. & Langridge, R. (1988). *J. Mol. Graph.* **6**, 13–27.
Gainey, L. D. S., Connerton, I. F., Lewis, E. H., Turner, G. & Ballance, D. J. (1992). *Curr. Gen.* **21**, 43–47.
Helin, S., Kahn, P. C., Guha, B. L. Mallows, D. G. & Goldman, A. (1995). *J. Mol. Biol.* **254**, 918–941.
Huang, C. C., Pettersen, E. F., Klein, T. E., Ferrin, T. E. & Langridge, R. (1991). *J. Mol. Graph.* **9**, 230–236.
Huang, K., Li, Z., Jia, Y., Dunaway-Mariano, D. & Herzberg, O. (1999). *Structure*, **7**, 539–548.
Jones, T. A. (1985). *Methods Enzymol.* **115**, 157–171.
Kabsch, W. & Sander, C. (1983). *Biopolymers*, **22**, 2577–2637.
Ko, Y. H., Cremo, C. R. & McFadden, B. A. (1992). *J. Biol. Chem.* **267**, 91–95.
Kornberg, H. L. & Krebs, H. A. (1957). *Nature (London)*, **179**, 988–991.
Kraulis, P. J. (1991). *J. Appl. Cryst.* **24**, 946–950.
Lamzin, V. S. & Wilson, K. S. (1993). *Acta Cryst.* D**49**, 129–147.
Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.
Lee, B. & Richards, F. M. (1971). *J. Mol. Biol.* **55**, 379–400.
Leslie, A. G. W. (1992). *Jnt CCP4/ESF–EACBM Newsl. Protein Crystallogr.* **26**.
McFadden, B. A. & Purohit, S. (1977). *J. Bacteriol.* **131**, 136–144.
Matsuoka, M. & McFadden, B. A. (1988). *J. Bacteriol.* **170**, 4528–4536.
Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). *Acta Cryst.* D**53**, 240–255.
Neidhart, D. J., Howell, P. L., Petsko, G. A., Powers, V. M., Li, R., Kenyon, G. L. & Gerlt, J. A. (1991). *Biochemistry*, **30**, 9264–9273.
Rehman, A. & McFadden, B. A. (1997). *Curr. Microbiol.* **34**, 205–211.
Schloss, J. V. & Cleland, W. W. (1982). *Biochemistry*, **21**, 4420–4427.
Sharma, V., Sharma, S., Hoener zu Bentrup, K., McKinney, J. D., Russell, D. G., Jacobs, W. R. & Sacchettini, J. C. (2000). *Nature Struct. Biol.* **7**, 663–668.
Shimamoto, G. & Berk, R. S. (1980). *Biochim. Biophys. Acta*, **632**, 399–407.
Simon, M. W., Martin, E. & Mukkada, A. J. (1978). *J. Bacteriol.* **135**, 895–899.
Vanni, P., Giaghetti, E., Pinzauti, G. & McFadden, B. A. (1990). *Comput. Biochem. Physiol.* **95**B, 431–458.